

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C07K 14/00</b>	<b>A2</b>	(11) International Publication Number: <b>WO 99/00410</b> (43) International Publication Date: <b>7 January 1999 (07.01.99)</b>
(21) International Application Number: <b>PCT/US98/13012</b> (22) International Filing Date: <b>23 June 1998 (23.06.98)</b> (30) Priority Data: <b>08/884,072</b> <b>27 June 1997 (27.06.97)</b> <b>US</b> (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application <b>08/884,072 (CIP)</b> <b>US</b> <b>Filed on 27 June 1997 (27.06.97)</b> (71) Applicant (for all designated States except US): <b>INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).</b> (72) Inventors; and (75) Inventors/Applicants (for US only): <b>BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US).</b>		(74) Agent: <b>BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).</b> (81) Designated States: <b>AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>

Published

*Without international search report and to be republished upon receipt of that report.*

(54) Title: **HUMAN EXTRACELLULAR MATRIX PROTEINS**

```
          9          18          27          36          45          54
CCA AGA TTG TTG TGA GGA GTC TAG CCA GTT GGT GAG CGC TGT AAT CTG AAC CAG

          63          72          81          90          99          108
CTG TGT CCA GAC TGA GGC CCC ATT TGC ATT ATT TAA CAT ACT TAG AAA ATG AAG

          117         126         135         144         153         162
TGT TCA TTT TTA ACA TTC CTC CTC CAA TTG GTT TAA TGC TGA AIT ACT GAA GAG

          171         180         189         198         207         216
GGC TAA GCA AAA CCA GGT GCT TGC GCT GAG GGC TCT GCA GTG GCT GGG AGG ACC

          225         234         243         252         261         270
CCG GCG CTC TCC CCG TGT CCT CTC CAC GAC TCG CTC GGC CCC TCT GGA ATA AAA

          279         288         297         306         315         324
CAC CCG CGA GCC CCG AGG GCC CAG AGG AGG CCG ACG TGC CCG AGC TCC TCC GGG

          333         342         351         360         369         378
GGT CCC GCC CGC GAG CTT TCT TCT CGC CTT CGC ATC TCC TCC TCG CGC GTC TTG

          387         396         405         414         423         432
GAC ATG CCA GGA ATA AAA AGG ATA CTC ACT GTT ACC AIT CTG GCT CTC TGT CTT
M   P   G   I   K   R   I   L   T   V   T   I   L   A   L   C   L
```

(57) Abstract

The invention provides two human extracellular matrix proteins (ECMP) and polynucleotides which identify and encode ECMP. The invention also provides expression vectors, host cells, agonists, antibodies and antagonists. The invention also provides methods for treating disorders associated with expression of ECMP.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## HUMAN EXTRACELLULAR MATRIX PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human extracellular matrix proteins and to the use of these sequences in the diagnosis, prevention, and treatment of cancer and immune disorders.

5

### BACKGROUND OF THE INVENTION

Many eukaryotic cells are enveloped by an extracellular matrix of proteins that provide structural support, cell and tissue identity, and autocrine, paracrine and juxtacrine properties for the cell within its environment (McGowan, S.E. (1992) *FASEB J.* 6:2895-2904). The diverse  
10 biochemistry of extracellular matrix proteins (ECMP) is indicative of the many, often overlapping, roles that are attributed to each distinct molecule (cf. Grant, D.S. and Kleinman, H.K. (1997) *E.X.S.* 79:317-333). Whilst a great number of ECMPs have been isolated, it still remains unclear how the majority interact with other ECMPs or with molecules residing within the cell membrane. Many ECMPs have been associated with tissue growth and cell proliferation,  
15 others with tissue or cell differentiation, and yet others with cell death (cf. Taipale, J. and Keski-Oja, J. (1997) *FASEB J.* 11:51-59; Eleftheriou, C.S. et al. (1991) *Mutat. Res.* 256:127-138).

For example, the process of embryonic bone formation involves the creation of an extracellular matrix that mineralizes during the course of tissue maturation. During the life of an individual, this matrix is subject to constant remodeling, through the combined actions of  
20 osteoblasts (which form mineralized bone) and osteoclasts (which resorb bone). The balance of ECMP composition, and the resulting bone structure, may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases (Francomano, C.A. et al. (1996) *Curr. Opin. Genet. Dev.* 6:301-308).

ECMPs also act as important mediators and regulators during the inflammatory response.  
25 Leukocytes are primed for inflammatory mediator and cytokine production by binding to ECMPs during extravasation (Pakianathan, D.R. (1995) *J. Leukoc Biol.* 57:699-702). Deposition of ECMPs is also triggered by inflammation in response to lung injury (Roman, J. (1996) *Immunol. Res.* 15:163-178). Although the function of newly deposited matrices in injured lungs is unknown, their ability to affect the migration, proliferation, differentiation, and activation state of

cells in vitro suggested an important role in the initiation and maintenance of the inflammatory response in vivo (Roman, J. supra)

Some examples of recently identified ECMPs which regulate cellular and tissue differentiation are S1-5 and *Ecm1*. S1-5 mRNA is overexpressed both in senescent human  
5 fibroblasts established from a subject with Werner syndrome of premature ageing and in growth-arrested normal human fibroblasts (Lecka-Czernik, B. et al. (1995) Mol. Cell. Biol. 15:120-128). The mRNA encodes a 387 amino acid residue protein containing five epidermal growth factor (EGF)-like domains. These domains matched the EGF tandem repeat consensus within several  
10 known extracellular proteins that promote cell growth, development, and cell signaling. The EGF tandem repeat is characterized by a regular distribution of single cysteines. As occurs with other members of the EGF-like family, the S1-5 gene product may represent a negative and/or positive factor whose ultimate activity is modulated by the cell environment (Lecka-Czernik, B. supra).

Murine *Ecm1* encodes a 559 residue protein that has been localized to one genetic locus associated with developmental disorders of the skin (Bhalerao, J. et al. (1995) J. Biol. Chem.  
15 270:16385-16394). During embryonic development, the gene is predominantly expressed in the form of splice variants in skin or cartilage tissue. Expression of the *Ecm1* gene also peaks during the late, pre-confluence phase of the murine osteogenic cell line, MN7, which proliferates and differentiates in vitro forming a mineralized matrix (Bhalerao, J. et al. (1995) supra). The murine  
20 *Ecm1* gene has been localized by genetic mapping to mouse chromosome 3, a region homologous to that of human chromosome 1q21 (Bhalerao, J. et al. (1995) supra). The molecular structure of the predicted protein is characterized by a pair of domains which share internal homology, and by a regular distribution of single cysteines and cysteine doublets. The latter arrangement was  
25 predicted to generate characteristic 'double-loop' proteins in the serum albumin family of proteins (Soltysik-Espanola, M. et al. (1994) Dev. Biol. 165:73-85). These double-loop' structures are involved in important ligand-binding functions (Kragh-Hansen, U. (1990) Danish Med. Bull. 37:57-84).

The discovery of new human extracellular matrix proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention and treatment of cancer and immune disorders.

## SUMMARY OF THE INVENTION

The invention features substantially purified human extracellular matrix proteins.

collectively referred to as ECMP and individually as ECMP-1, and ECMP-2, having the amino acid sequence shown in SEQ ID NO:1, or SEQ ID NO:3, respectively, or fragments thereof.

The invention further provides an isolated and substantially purified polynucleotide sequence encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof and a composition comprising said polynucleotide sequence. The invention  
5 also provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence encoding the amino acid sequence SEQ ID NO:1, or fragments of said polynucleotide sequence. The invention further provides a polynucleotide sequence comprising the complement of the polynucleotide sequence encoding the amino acid sequence of SEQ ID  
10 NO:1, or fragments or variants of said polynucleotide sequence.

The invention also provides an isolated and purified sequence comprising SEQ ID NO:2 or variants thereof. In addition, the invention provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of SEQ ID NO:2. In another aspect the invention provides a composition comprising an isolated and purified  
15 polynucleotide sequence comprising the complement of SEQ ID NO:2, or fragments or variants thereof. The invention also provides a polynucleotide sequence comprising the complement of SEQ ID NO:2.

The present invention further provides an expression vector containing at least a fragment of any of the claimed polynucleotide sequences. In yet another aspect, the expression vector  
20 containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment thereof, the method comprising the steps of: a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding ECMP-1 under conditions suitable for the expression of the  
25 polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified ECMP-1 having the amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

The invention also provides a purified antagonist which decreases the activity of a  
30 polypeptide of SEQ ID NO:1. In one aspect the invention provides a purified antibody which binds to a polypeptide comprising at least a fragment of the amino acid sequence of SEQ ID NO:1.

Still further, the invention provides a purified agonist which modulates the activity of the polypeptide of SEQ ID NO:1.

The invention also provides a method for treating or preventing cancer comprising administering to a subject in need of such treatment an effective amount of a purified antagonist of ECMP-1.

The invention also provides a method for treating or preventing an immune disorder comprising administering to a subject in need of such treatment an effective amount of a purified antagonist of ECMP-1.

The invention also provides a method for detecting a polynucleotide which encodes ECMP-1 in a biological sample comprising the steps of: a) hybridizing a polynucleotide sequence complementary to the polynucleotide sequence encoding ECMP-1 (SEQ ID NO:1) to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding ECMP-1 in the biological sample. In a preferred embodiment, prior to hybridization, the nucleic acid material of the biological sample is amplified by the polymerase chain reaction.

The invention further provides an isolated and substantially purified polynucleotide sequence encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:3 or fragments thereof and a composition comprising said polynucleotide sequence. The invention also provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence encoding the amino acid sequence SEQ ID NO:3, or fragments of said polynucleotide sequence. The invention further provides a polynucleotide sequence comprising the complement of the polynucleotide sequence encoding the amino acid sequence of SEQ ID NO:3, or fragments or variants of said polynucleotide sequence.

The invention also provides an isolated and purified sequence comprising SEQ ID NO:4 or variants thereof. In addition, the invention provides a polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of SEQ ID NO:4. In another aspect the invention provides a composition comprising an isolated and purified polynucleotide sequence comprising the complement of SEQ ID NO:4, or fragments or variants thereof. The invention also provides a polynucleotide sequence comprising the complement of SEQ ID NO:4.

The present invention further provides an expression vector containing at least a fragment

of any of the claimed polynucleotide sequences. In yet another aspect, the expression vector containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:3 or a fragment thereof, the method comprising the steps of: a) 5 culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding ECMP-2 under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified ECMP-2 having the amino acid sequence of SEQ ID NO:3 in conjunction with a suitable 10 pharmaceutical carrier.

The invention also provides a purified antagonist which decreases the activity of a polypeptide of SEQ ID NO:3. In one aspect the invention provides a purified antibody which binds to a polypeptide comprising at least a fragment of the amino acid sequence of SEQ ID NO:3.

15 Still further, the invention provides a purified agonist which modulates the activity of the polypeptide of SEQ ID NO:3.

The invention also provides a method for treating or preventing cancer comprising administering to a subject in need of such treatment an effective amount of a purified antagonist of ECMP-2.

20 The invention also provides a method for treating or preventing an immune disorder comprising administering to a subject in need of such treatment an effective amount of a purified antagonist of ECMP-2.

The invention also provides a method for detecting a polynucleotide which encodes ECMP-2 in a biological sample comprising the steps of: a) hybridizing a polynucleotide 25 sequence complementary to the polynucleotide sequence encoding ECMP-2 (SEQ ID NO:3) to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding ECMP-2 in the biological sample. In a preferred embodiment, the nucleic acid material of the biological sample is amplified by the polymerase 30 chain reaction prior to hybridization.

## BRIEF DESCRIPTION OF THE FIGURES

Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of ECMP-1. The alignment was produced using MacDNASIS PRO™ software (Hitachi Software Engineering Co. Ltd. San Bruno, CA).

5        Figures 2A, 2B, 2C, 2D, and 2E show the amino acid sequence (SEQ ID NO:3) and nucleic acid sequence (SEQ ID NO:4) of ECMP-2. The alignment was produced using MacDNASIS PRO™ software.

10        Figures 3A, and 3B show the amino acid sequence alignments between ECMP-1 (SEQ ID NO:1) and human S1-5 gene product (GI 458228; SEQ ID NO:5), produced using the multisequence alignment program of DNASTAR™ software (DNASTAR Inc, Madison WI).

      Figures 4A, 4B, and 4C show the amino acid sequence alignments between ECMP-2 (SEQ ID NO:3), and murine secreted protein encoded by *Ecml* gene (GI 496120; SEQ ID NO:6), produced using the multisequence alignment program of DNASTAR™ software.

## DESCRIPTION OF THE INVENTION

15        Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

20        It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

30        Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies



which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## 5 DEFINITIONS

ECMP, as used herein, refers to the amino acid sequences of substantially purified ECMP obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

10 The term "agonist", as used herein, refers to a molecule which, when bound to ECMP, increases or prolongs the duration of the effect of ECMP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of ECMP.

An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding ECMP. Alleles may result from at least one mutation in the nucleic acid sequence and  
15 may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

20 "Altered" nucleic acid sequences encoding ECMP as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent ECMP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ECMP, and improper or unexpected hybridization to  
25 alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding ECMP. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ECMP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the  
30 residues as long as the biological or immunological activity of ECMP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups

having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules.

5 Fragments of ECMP are preferably about 5 to about 15 amino acids in length and retain the biological activity or the immunological activity of ECMP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

10 "Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "antagonist" as used herein, refers to a molecule which, when bound to ECMP,  
15 decreases the amount or the duration of the effect of the biological or immunological activity of ECMP. Antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules which decrease the effect of ECMP.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, F(ab')<sub>2</sub>, and Fv, which are capable of binding the epitopic determinant.

20 Antibodies that bind ECMP polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet  
25 hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the  
30 production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune

response) for binding to an antibody.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand.

5 Antisense molecules include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

10 The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic ECMP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

15 The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the  
20 single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

A "composition comprising a given polynucleotide sequence" as used herein refers  
25 broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding ECMP (SEQ ID NO:1, SEQ ID NO:3 ) or fragments thereof (e.g., SEQ ID NO:2, or SEQ ID NO:4 and fragments thereof) may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as  
30 a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly (e.g., GELVIEW™ Fragment Assembly system, GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:2, or SEQ ID NO:4 by northern analysis is indicative of the presence of mRNA encoding ECMP in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to ECMP or the encoded ECMP. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the

probe will not hybridize to the second non-complementary target sequence.

Human artificial chromosomes (HACs) are linear microchromosomes which may contain DNA sequences of 10K to 10M in size and contain all of the elements required for stable mitotic chromosome segregation and maintenance (Harrington, J.J. et al. (1997) Nat Genet. 15:345-355).

5 The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

10 The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in  
15 solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

An "insertion" or "addition", as used herein, refers to a change in an amino acid or  
20 nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

"Microarray" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

25 The term "modulate", as used herein, refers to a change in the activity of ECMP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of ECMP.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which  
30 may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" are those nucleic acid sequences which are greater than 60 nucleotides in length, and most preferably includes fragments that are at least 100 nucleotides or at least 1000 nucleotides, and at

least 10,000 nucleotides in length.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or hybridization assays. As  
5 used herein, oligonucleotide is substantially equivalent to the terms "amplimers", "primers", "oligomers", and "probes", as commonly defined in the art.

"Peptide nucleic acid", PNA as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in lysine. The terminal lysine confers solubility to  
10 the composition. PNAs may be pegylated to extend their lifespan in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino  
15 acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:1, SEQ ID NO:3," encompasses the full-length ECMP and fragments thereof.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding ECMP, or fragments thereof, or ECMP itself may  
20 comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane isolated from a cell, a cell, genomic DNA, RNA, or cDNA (in solution or bound to a solid support, a tissue, a tissue print, and the like.

The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The  
25 interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) of the protein recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

30 The terms "stringent conditions" or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related

polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of ECMP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

## THE INVENTION

The invention is based on the discovery of a new human extracellular matrix proteins (collectively referred to as "ECMP" and individually, as ECMP-1 and ECMP-2), the polynucleotides encoding ECMP, and the use of these compositions for the diagnosis, prevention, or treatment of cancer and immune disorders.

Nucleic acids encoding the ECMP-1 of the present invention were first identified in Incyte Clone 45517 from the corneal fibroblast cDNA library (CORNNOT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 45517 (CORNNOT01), 424333 (BLADNOT01), 1322651 (BLADNOT04), 198548 (KIDNNOT02), 944281 (ADRENOT03), and 953977 (SCORNON01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G. ECMP-1 is 449 amino acids in length and has five potential EGF-like tandem repeats between C-113 and C-341, four of which have the consensus repeat of  $Cx\{11-12\}Cx\{5\}Cx\{4-6\}Cx\{3 \text{ or } 5\}Cx\{8\}Cx$  (where x represents any amino acid and {n} represents the number of residues). Within the EGF-like domain, there are four potential N-hydroxylation sites at N-146, N-183, N-223, and N-264; and two potential N-glycosylation sites, at N-283, and N-296. The N-terminal 19 amino acids constitute a potential signal peptide sequence, and there is a potential cell attachment sequence, RGD, at R-54. There are eight potential casein kinase II phosphorylation sites at T-48, S-154, T-197, S-204, S-246, S-252, S-285, and S-385, and two potential protein kinase C phosphorylation sites at S-357, and at T-371. There is a potential prokaryotic membrane lipoprotein attachment site between L-180 and C-190. As shown in Figures 3A and 3B, ECMP-1 has chemical and structural homology with human S1-5 gene product (GI 458228; SEQ ID NO:5). In particular, ECMP-1 shares 53% identity, an EGF-like tandem repeat motif, and potential N-hydroxylation sites with human S1-5 gene product. Northern analysis shows the expression of this sequence in various libraries, at least 20% of which are immortalized or cancerous and at least 28% of which involve the immune response.

Nucleic acids encoding the ECMP-2 of the present invention were first identified in Incyte Clone 1621777 from the brain tumor tissue cDNA library (BRAITUT13) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:4, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 1621777



(BRAITUT13), 865787 (BRAITUT03), 1867044 (SKINBIT01), 1901493 (BLADTUT06) and 1957753 (CONNNOT01).

In another embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:3, as shown in Figures 2A, 2B, 2C, 2D, and 2E. ECMP-2 is 540 amino acids in length and has six potential cysteine repeats of single cysteines and cysteine doublets between C-181 and C-498. These repeats are characteristic of domains in the serum albumin family of proteins (Soltysik-Espanola, M. et al. (1994) supra). There is a potential N-terminal signal peptide between M-1 and A-20 and two internal homology domains between C-151 to Y-279 and C-284 to Y-405. In addition, ECMP-2 has three potential N-glycosylation sites at N-354, N-444, and N-530, five potential casein kinase II phosphorylation sites at residues S-138, S-293, T-391, S-490, and S-533, five potential protein kinase C phosphorylation sites at T-4, T-227, S-250, T-358, and T-446, and one potential tyrosine kinase phosphorylation site at Y-374.

As shown in Figures 4A, 4B, and 4C, ECMP-2 has chemical and structural homology with the secreted protein encoded by the murine *Ecml* gene (GI 496120; SEQ ID NO:6). In particular, ECMP-2 shares 84% identity with mouse secreted protein. Both proteins share the single cysteine and cysteine doublet repeat domains. They share two of the potential N-glycosylation sites, three of the potential casein kinase II sites, and four of the potential protein kinase C sites. They also share the two internal sequence homology domains of *Ecml*. Northern analysis shows the expression of this sequence in various libraries, at least 32% of which are immortalized or cancerous and at least 39% of which involve the immune response.

The invention also encompasses ECMP variants. A preferred ECMP variant is one having at least 80%, and more preferably 90%, amino acid sequence identity to the ECMP amino acid sequence (SEQ ID NO:1, or SEQ ID NO:3) and which retains at least one of the biological, structural or other functional characteristics of ECMP. A most preferred ECMP variant is one having at least 95% amino acid sequence identity to SEQ ID NO:1, or SEQ ID NO:3.

The invention also encompasses polynucleotides which encode ECMP. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of ECMP can be used to produce recombinant molecules which express ECMP. In a particular embodiment, the invention encompasses the polynucleotide comprising the nucleic acid sequence of SEQ ID NO:2, or SEQ ID NO:4 as shown in Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G, Figures 2A, 2B, 2C, 2D, and 2E, respectively.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding ECMP, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide  
5 sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring ECMP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ECMP and its variants are preferably  
10 capable of hybridizing to the nucleotide sequence of the naturally occurring ECMP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ECMP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular  
15 codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ECMP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof,  
20 which encode ECMP and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ECMP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of  
25 hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:2, or SEQ ID NO:4, under various conditions of stringency as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and  
30 may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham,

Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD).

Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the  
5 ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding ECMP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent  
10 to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse  
15 transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30  
20 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of  
25 DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

30 Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process

avoids the need to screen libraries and is useful in finding intron/exon junctions. When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode ECMP may be used in recombinant DNA molecules to direct expression of ECMP, fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express ECMP.

As will be understood by those of skill in the art, it may be advantageous to produce ECMP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ECMP encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene

fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding ECMP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of ECMP activity, it may be useful to encode a chimeric ECMP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the ECMP encoding sequence and the heterologous protein sequence, so that ECMP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding ECMP may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of ECMP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of ECMP, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active ECMP, the nucleotide sequences encoding ECMP or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding ECMP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques,

synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

5 A variety of expression vector/host systems may be utilized to contain and express sequences encoding ECMP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression  
10 vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular  
15 proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid  
20 (Gibco/BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell  
25 line that contains multiple copies of the sequence encoding ECMP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for ECMP. For example, when large quantities of ECMP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are  
30 readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding ECMP may be ligated into the vector in frame with sequences for the amino-terminal

Met and the subsequent 7 residues of  $\beta$ -galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding ECMP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express ECMP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding ECMP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of ECMP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which ECMP may be expressed (Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In

cases where an adenovirus is used as an expression vector, sequences encoding ECMP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing ECMP in infected host cells  
5 (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10M are constructed  
10 and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding ECMP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding ECMP, its initiation codon, and upstream  
15 sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation  
20 codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of  
25 the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic  
30 mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.



For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express ECMP may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk<sup>-</sup> or apt<sup>-</sup> cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins,  $\beta$  glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding ECMP is inserted within a marker gene sequence, transformed cells containing sequences encoding ECMP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ECMP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding ECMP and express ECMP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding ECMP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding ECMP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding ECMP to detect transformants containing DNA or RNA encoding ECMP.

A variety of protocols for detecting and measuring the expression of ECMP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on ECMP is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ECMP include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ECMP, or any fragments thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding ECMP may be cultured under

conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ECMP may be designed to contain signal sequences which direct secretion of ECMP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding ECMP to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and ECMP may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing ECMP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying ECMP from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

In addition to recombinant production, fragments of ECMP may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of ECMP may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

## THERAPEUTICS

Chemical and structural homology exists among ECMP-1 and human S1-5 gene product (GI 458228). In addition, ECMP-1 is expressed in tissues associated with cancer and the immune response. Therefore, ECMP-1 appears to play a role in cancer and immune disorders, particularly disorders in which ECMP-1 is overexpressed.

Therefore, in one embodiment, an antagonist of ECMP-1 may be administered to a subject to prevent or treat cancer. Cancers may include, but are not limited to adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma, and particularly cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector expressing the complement of the polynucleotide encoding ECMP-1 may be administered to a subject to treat or prevent cancer including, but not limited to, the types of cancer described above.

In another embodiment, an antagonist of ECMP-1 may be administered to a subject to prevent or treat an immune disorder. Such disorders may include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, Werner syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma.

In another embodiment, a vector expressing the complement of the polynucleotide encoding ECMP-1 may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those described above.

In one aspect, antibodies which specifically bind ECMP-1 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express ECMP-1.

Chemical and structural homology exists among ECMP-2 and murine secreted protein encoded by *Ecml* gene (GI 496120). In addition, ECMP-2 is expressed in tissues associated with cancer and the immune response. Therefore, ECMP-2 appears to play a role in cancer and immune disorders, particularly disorders in which ECMP-2 is overexpressed.

Therefore, in another embodiment, an antagonist of ECMP-2 may be administered to a subject to prevent or treat cancer. Cancers may include, but are not limited to, adenocarcinoma,

leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma, and particularly cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

5           In another embodiment, a vector expressing the complement of the polynucleotide encoding ECMP-2 may be administered to a subject to treat or prevent a cancer including, but not limited to, any of the types of cancer described above.

          In another embodiment, an antagonist of ECMP-2 may be administered to a subject to prevent or treat an immune disorder. Such disorders may include, but are not limited to, AIDS,  
10   Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,  
15   osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, Werner syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma.

          In another embodiment, a vector expressing the complement of the polynucleotide  
20   encoding ECMP-2 may be administered to a subject to treat or prevent an immune disorder including, but not limited to, those described above.

          In one aspect, an antibody which specifically binds ECMP-2 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express ECMP-2.

25           In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment  
30   or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of ECMP may be produced using methods which are generally known in the art. In particular, purified ECMP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ECMP.

Antibodies to ECMP may be generated using methods that are well known in the art.

- 5 Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with ECMP or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and  
15 Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ECMP have an amino acid sequence consisting of at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small,  
20 naturally occurring molecule. Short stretches of ECMP amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to ECMP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are  
25 not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing  
30 of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature

314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ECMP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for ECMP may also be generated. For example, such fragments include, but are not limited to, the F(ab')<sub>2</sub> fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ECMP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ECMP epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding ECMP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding ECMP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding ECMP. Thus, complementary molecules or fragments may be used to modulate ECMP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding ECMP.

Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or

from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding ECMP. These techniques are described both in

5 Sambrook et al. (supra) and in Ausubel et al. (supra).

Genes encoding ECMP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes ECMP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA  
10 molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or  
15 regulatory regions of the gene encoding ECMP (signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription  
20 factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

25 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding ECMP.

30 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20



ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

5 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding ECMP. Such DNA sequences may be  
10 incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'  
15 ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by  
20 endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino  
25 polymers (Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

30 An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of

ECMP, antibodies to ECMP, mimetics, agonists, antagonists, or inhibitors of ECMP. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The  
5 compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal,  
10 intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found  
15 in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets,  
20 pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores.  
25 Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl  
30 pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel,

polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of

ECMP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ECMP or fragments thereof, antibodies of ECMP, agonists, antagonists or inhibitors of ECMP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and

methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

5

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind ECMP may be used for the diagnosis of conditions or disorders characterized by expression of ECMP, or in assays to monitor patients being treated with ECMP, agonists, antagonists or inhibitors. The antibodies useful for  
10 diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for ECMP include methods which utilize the antibody and a label to detect ECMP in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in  
15 the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring ECMP are known in the art and provide a basis for diagnosing altered or abnormal levels of ECMP expression. Normal or standard values for ECMP expression are established by combining body fluids or cell  
20 extracts taken from normal mammalian subjects, preferably human, with antibody to ECMP under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of ECMP expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the  
parameters for diagnosing disease.

25 In another embodiment of the invention, the polynucleotides encoding ECMP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of ECMP may be correlated with disease. The diagnostic assay may be used to distinguish between  
30 absence, presence, and excess expression of ECMP, and to monitor regulation of ECMP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting

polynucleotide sequences, including genomic sequences, encoding ECMP or closely related molecules, may be used to identify nucleic acid sequences which encode ECMP. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding ECMP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the ECMP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:2, or SEQ ID NO:4 or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring ECMP.

Means for producing specific hybridization probes for DNAs encoding ECMP include the cloning of nucleic acid sequences encoding ECMP or ECMP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding ECMP may be used for the diagnosis of conditions, or disorders which are associated with expression of ECMP. Examples of such conditions or diseases include cancer such as cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, Werner syndrome, and autoimmune thyroiditis;

complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma. The polynucleotide sequences encoding ECMP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays or  
5 microarrays utilizing fluids or tissues from patient biopsies to detect altered ECMP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ECMP may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding ECMP may be labeled by standard methods, and added to a fluid  
10 or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of  
15 nucleotide sequences encoding ECMP in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of  
20 ECMP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes ECMP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a  
25 substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to  
30 approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ECMP may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'→3') and another with antisense (3'←5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of ECMP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides derived from any of the polynucleotide sequences described herein may be used as targets in microarrays. The microarrays can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information will be useful in determining gene function, understanding the genetic basis of disease, diagnosing disease, and in developing and monitoring the activity of therapeutic agents (Heller, R. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-55).

In one embodiment, the microarray is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference.



The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray, it may be preferable to use oligonucleotides which are only 7-10 nucleotides in length. The microarray may contain oligonucleotides which cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell type, developmental or disease state.

In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' or more preferably at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray. The "pairs" will be identical, except for one nucleotide which preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, the oligomers may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array may be produced by hand or using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments) and may contain 8, 24, 96,

384, 1536 or 6144 oligonucleotides, or any other multiple between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarrays, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray so that the probe sequences hybridize to complementary oligonucleotides of the microarray. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, mutations, variants, or polymorphisms among samples.

In another embodiment of the invention, the nucleic acid sequences which encode ECMP may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C.M. (1993) *Blood Rev.* 7:127-134, and Trask, B.J. (1991) *Trends Genet.* 7:149-154.

Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at Online Mendelian Inheritance in Man (OMIM). Correlation between the location of the gene encoding ECMP on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the

subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, ECMP, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between ECMP and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to ECMP large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with ECMP, or fragments thereof, and washed.

Bound ECMP is then detected by methods well known in the art. Purified ECMP can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ECMP specifically compete with a test compound for binding ECMP. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ECMP.

In additional embodiments, the nucleotide sequences which encode ECMP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

5       The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

## EXAMPLES

### I       cDNA Library Construction

#### 10    CORNNOT01

The corneal fibroblast CORNNOT01 cDNA library was custom constructed by Stratagene using stromal RNA isolated from the corneal fibroblasts of a 76-year-old. Stratagene prepared the cDNA library using an XhoI-oligo d(T) primer. Double-stranded cDNA was blunted, ligated to EcoRI adaptors, digested with XhoI, size-selected, and cloned into the XhoI and EcoRI sites of the Lambda UniZAP™ vector (Stratagene). Following Lambda UniZAP™ packaging, 2 x 10<sup>6</sup> primary clones were then amplified to stabilize the library for long-term storage.

The quality of the cDNA library was screened using DNA probes, and then, the Bluescript® phagemid (Stratagene) was excised. Subsequently, the custom-constructed library phage particles were infected into E. coli host strain XL1-Blue® (Stratagene). Alternative  
20   unidirectional vectors include, but are not limited to, pcDNA1 (Invitrogen, Carlsbad, CA) and pSHlox-1 (Novagen, Madison WI).

#### BRAITUT13

The brain tumor BRAITUT13 cDNA library was constructed from cancerous brain tissue  
25   obtained from a 68-year-old Caucasian male during cerebral meningeal excision following diagnosis of meningioma localized in the left frontal part of the brain. In a prior surgery the patient had undergone a replacement of aortic valve with tissue graft.

The frozen tissue was homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The  
30   lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.7, precipitated using 0.3 M sodium

acetate and 2.5 volumes of ethanol, resuspended in RNase-free water, and DNase treated at 37°C. Extraction and precipitation were repeated as before. The mRNA was then isolated with the Qiagen Oligotex kit (QIAGEN, Inc.; Chatsworth, CA) and used to construct the cDNA library.

5 The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA Synthesis and Plasmid Cloning (Cat. #18248-013; Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into pSport I. The plasmid pSport I was subsequently transformed into DH5 $\alpha$ <sup>TM</sup> competent cells (Cat. #18258-012; Gibco/BRL).

## 10 II Isolation and Sequencing of cDNA Clones

### CORNNOT01

The phagemid forms of individual cDNA clones were obtained by the in vivo excision process, in which the host bacterial strain, XPORT E. coli (Stratagene, XL1 Blue strain) was  
15 coinfectd with both the lambda library phage (Lambda ZAP<sup>®</sup>II, Stratagene) and an f1 helper phage (Ex Assist, Stratagene). Polypeptides or enzymes derived from both the library-containing phage and the helper phage nicked the DNA, initiating new DNA synthesis from defined sequences on the target DNA and creating a smaller, single stranded circular phagemid DNA molecule that included all DNA, sequences of the pBluescript phagemid and the cDNA insert.  
20 The phagemid DNA was released from the cells and purified, then used to re-infect fresh host cells (SOLR, Stratagene) where the double stranded DNA was produced. Because the phagemid carries the gene for  $\beta$ -lactamase, the newly-transformed bacteria were selected on medium containing ampicillin.

Phagemid DNA was purified using the QIAwell<sup>TM</sup>-8 Plasmid Purification System,  
25 QIAwell PLUS, or QIAwell ULTRA DNA purification system (QIAGEN, Inc., Chatsworth, CA). An alternative method for purifying the phagemid utilizes the Miniprep Kit available from Advanced Genetic Technologies Corp. (Gaithersburg, MD).

### BRAITUT13

30 Plasmid DNA was released from the cells and purified using the REAL Prep 96 Plasmid Kit (Catalog #26173; QIAGEN, Inc.). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711,

Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems.

### III Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences of the Sequence Listing or amino acid sequences deduced from them were used as query sequences against databases such as GenBank, SwissProt, BLOCKS, and Pima II. These databases which contain previously identified and annotated sequences were searched for regions of homology (similarity) using BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul et al. (1990) J. Mol. Biol. 215:403-410).

BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal or plant) origin. Other algorithms such as the one described in Smith R.F. and T.F. Smith (1992; Protein Engineering 5:35-51), incorporated herein by reference, can be used when dealing with primary sequence patterns and secondary structure gap penalties. As disclosed in this application, the sequences have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach, as detailed in Karlin, S. and S.F. Atschul (1993; Proc. Nat. Acad. Sci. 90:5873-7) and incorporated herein by reference, searches for matches between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. In this application, threshold was set at  $10^{-25}$  for nucleotides and  $10^{-14}$  for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate (pri), rodent (rod), and mammalian sequences (mam), and deduced amino acid sequences from

the same clones are searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp) and eukaryote (eukp), for homology. The relevant database for a particular match were reported as a Gxxxx±p (where xxx is pri, rod, etc and if present, p = peptide).

#### 5 IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., supra).

10 Analogous computer techniques using BLAST (Altschul, S.F. 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$15 \quad \frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually  
20 identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding ECMP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and  
25 percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

#### V Extension of ECMP Encoding Polynucleotides

30 The nucleic acid sequence of the Incyte Clone 45517 or 162177 was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension in the antisense direction, and the other was synthesized to extend sequence in the sense direction. Primers were used to facilitate the extension of the

known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be about 22 to about 30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures of about 68° to about 72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (Gibco/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13  $\mu$ l of ligation buffer, 1  $\mu$ l T4-DNA ligase (15 units) and 1  $\mu$ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent *E. coli* cells (in 40  $\mu$ l of appropriate media) were transformed with 3  $\mu$ l of ligation mixture and cultured in 80  $\mu$ l



of SOC medium (Sambrook et al., supra). After incubation for one hour at 37° C, the *E. coli* mixture was plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150  $\mu$ l of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-  
5 available, sterile 96-well microtiter plate. The following day, 5  $\mu$ l of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5  $\mu$ l of each sample was transferred into a PCR array.

For PCR amplification, 18  $\mu$ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for  
10 the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
15 Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

20 Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2, or SEQ ID NO:4, is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5'  
25 extension, and an appropriate genomic library.

## VI Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2, or SEQ ID NO:4 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides,  
30 consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-  
35 25 superfine resin column (Pharmacia & Upjohn). A aliquot containing 10<sup>7</sup> counts per minute of

the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to  
5 nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphorimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours,  
10 hybridization patterns are compared visually.

## VII Microarrays

To produce oligonucleotides for a microarray, the nucleotide sequence described herein is examined using a computer algorithm which starts at the 3' end of the nucleotide sequence.  
15 The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that would interfere with hybridization. The algorithm identifies 20 sequence-specific oligonucleotides of 20 nucleotides in length (20-mers). A matched set of oligonucleotides is created in which one nucleotide in the center of each sequence is altered. This process is repeated  
20 for each gene in the microarray, and double sets of twenty 20 mers are synthesized and arranged on the surface of the silicon chip using a light-directed chemical process (Chee, M. et al., PCT/WO95/11995, incorporated herein by reference).

In the alternative, a chemical coupling procedure and an ink jet device are used to synthesize oligomers on the surface of a substrate (Baldeschweiler, J.D. et al.,  
25 PCT/WO95/25116, incorporated herein by reference). In another alternative, a "gridded" array analogous to a dot (or slot) blot is used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array may be produced by hand or using available materials and machines and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots. After  
30 hybridization, the microarray is washed to remove nonhybridized probes, and a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide

sequence on the micro-array.

### VIII Complementary Polynucleotides

Sequence complementary to the ECMP-encoding sequence, or any part thereof, is used  
5 to decrease or inhibit expression of naturally occurring ECMP. Although use of oligonucleotides  
comprising from about 15 to about 30 base-pairs is described, essentially the same procedure is  
used with smaller or larger sequence fragments. Appropriate oligonucleotides are designed using  
Oligo 4.06 software and the coding sequence of ECMP, SEQ ID NO:1, or SEQ ID NO:3. To  
inhibit transcription, a complementary oligonucleotide is designed from the most unique 5'  
10 sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a  
complementary oligonucleotide is designed to prevent ribosomal binding to the ECMP-encoding  
transcript.

### IX Expression of ECMP

15 Expression of ECMP is accomplished by subcloning the cDNAs into appropriate vectors  
and transforming the vectors into host cells. In this case, the cloning vector is also used to  
express ECMP in *E. coli*. Upstream of the cloning site, this vector contains a promoter for  
 $\beta$ -galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent  
seven residues of  $\beta$ -galactosidase. Immediately following these eight residues is a bacteriophage  
20 promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods  
produces a fusion protein which consists of the first eight residues of  $\beta$ -galactosidase, about 5 to  
15 residues of linker, and the full length protein. The signal residues direct the secretion of  
ECMP into the bacterial growth media which can be used directly in the following assay for  
25 activity.

### X Demonstration of ECMP Activity

The activity of ECMP-1 and ECMP-2 may be measured using an assay based upon the  
property of ECMPs to support proliferation in vitro of fibroblasts and tumor cells under serum-  
30 free conditions (Chiquet-Ehrismann, R, et al. (1986) Cell 47:131-139). Wells in 96 well cluster  
plates (Falcon, Fisher Scientific, Santa Clara, CA) are coated with ECMP by incubation with  
solutions at 50-100  $\mu$ g/ml for 15 min at ambient temperature. The coating solution is aspirated,

and the wells washed with Dulbecco's medium before cells are plated. Rat fibroblast cultures or rat mammary tumor cells are prepared as described and plated at a density of  $10^4$ - $10^5$  cells/ml in Dulbecco's medium supplemented with 10% fetal calf serum.

After three days the media are removed, and the cells washed three times with phosphate-buffered saline (PBS) before the addition of serum-free Dulbecco's medium containing 0.25 mg/ml bovine serum albumin (BSA, Fraction V, Sigma Chemical, St. Louis, MO). After 2 days the medium is aspirated, and 100  $\mu$ l of [ $^3$ H]thymidine (NEN) at 2  $\mu$ Ci/ml in fresh Dulbecco's medium containing 0.25 mg/ml BSA added. Parallel plates are fixed and stained to determine cell numbers. After 16 hr, the medium is aspirated, the cell layer washed with PBS, and the 10% trichloroacetic acid-precipitable counts in the cell layer determined by liquid scintillation counting of radioisotope (normalized to relative cell numbers; Chiquet-Ehrismann, R. et al. (1986) supra).

## **XI Production of ECMP Specific Antibodies**

ECMP that is substantially purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from SEQ ID NO:2, or SEQ ID NO:4 is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (supra), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al., supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio iodinated, goat anti-rabbit IgG.

## **XII Purification of Naturally Occurring ECMP Using Specific Antibodies**

Naturally occurring or recombinant ECMP is substantially purified by immunoaffinity chromatography using antibodies specific for ECMP. An immunoaffinity column is constructed

by covalently coupling ECMP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing ECMP is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ECMP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ECMP binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ECMP is collected.

### **XIII Identification of Molecules Which Interact with ECMP**

ECMP or biologically active fragments thereof are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133: 529). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ECMP, washed and any wells with labeled ECMP complex are assayed. Data obtained using different concentrations of ECMP are used to calculate values for the number, affinity, and association of ECMP with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

- (i) APPLICANT: INCYTE PHARMACEUTICALS, INC.
- (ii) TITLE OF THE INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS
- (iii) NUMBER OF SEQUENCES: 6
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
  - (B) STREET: 3174 Porter Drive
  - (C) CITY: Palo Alto
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
  - (A) PCT APPLICATION NUMBER: To Be Assigned
  - (B) FILING DATE: Herewith
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) U. S. APPLICATION NUMBER: 08/884,072
  - (B) FILING DATE: 27 June 1997
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Billings, Lucy J.
  - (B) REGISTRATION NUMBER: 36,749
  - (C) REFERENCE/DOCKET NUMBER: PF-0333 PCT
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 650-855-0555
  - (B) TELEFAX: 650-845-4166
  - (C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 448 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: CORNNOT01
  - (B) CLONE: 45517

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

Met Pro Gly Ile Lys Arg Ile Leu Thr Val Thr Ile Leu Ala Leu Cys
 1             5             10             15
Leu Pro Ser Pro Gly Asn Ala Gln Ala Gln Cys Thr Asn Gly Phe Asp
                20             25             30
Leu Asp Arg Gln Ser Gly Gln Cys Leu Asp Ile Asp Glu Cys Arg Thr
                35             40             45
Ile Pro Glu Ala Cys Arg Gly Asp Met Met Cys Val Asn Gln Asn Gly

```

50		55		60
Gly Tyr Leu Cys Ile Pro Arg Thr Asn Pro Val Tyr Arg Gly Pro Tyr				
65		70		75
Ser Asn Pro Tyr Ser Thr Pro Tyr Ser Gly Pro Tyr Pro Ala Ala Ala				80
	85		90	
Pro Pro Leu Ser Ala Pro Asn Tyr Pro Thr Ile Ser Arg Pro Leu Ile				95
	100		105	
Cys Arg Phe Gly Tyr Gln Met Asp Glu Ser Asn Gln Cys Val Asp Val				110
	115		120	
Asp Glu Cys Ala Thr Asp Ser His Gln Cys Asn Pro Thr Gln Ile Cys				125
	130		135	
Ile Asn Thr Glu Gly Tyr Thr Cys Ser Cys Thr Asp Gly Tyr Trp				140
145		150		155
Leu Leu Glu Gly Gln Cys Leu Asp Ile Asp Glu Cys Arg Tyr Gly Tyr				160
	165		170	
Cys Gln Gln Leu Cys Ala Asn Val Pro Gly Ser Tyr Ser Cys Thr Cys				175
	180		185	
Asn Pro Gly Phe Thr Leu Asn Glu Asp Gly Arg Ser Cys Gln Asp Val				190
	195		200	
Asn Glu Cys Ala Thr Glu Asn Pro Cys Val Gln Thr Cys Val Asn Thr				205
	210		215	
Tyr Gly Ser Phe Ile Cys Arg Cys Asp Pro Gly Tyr Glu Leu Glu Glu				220
225		230		235
Asp Gly Val His Cys Ser Asp Met Asp Glu Cys Ser Phe Ser Glu Phe				240
	245		250	
Leu Cys Gln His Glu Cys Val Asn Gln Pro Gly Thr Tyr Phe Cys Ser				255
	260		265	
Cys Pro Pro Gly Tyr Ile Leu Leu Asp Asp Asn Arg Ser Cys Gln Asp				270
	275		280	
Ile Asn Glu Cys Glu His Arg Asn His Thr Cys Asn Leu Gln Gln Thr				285
	290		295	
Cys Tyr Asn Leu Gln Gly Gly Phe Lys Cys Ile Asp Pro Ile Arg Cys				300
305		310		315
Glu Glu Pro Tyr Leu Arg Ile Ser Asp Asn Arg Cys Met Cys Pro Ala				320
	325		330	
Glu Asn Pro Gly Cys Arg Asp Gln Pro Phe Thr Ile Leu Tyr Arg Asp				335
	340		345	
Met Asp Val Val Ser Gly Arg Ser Val Pro Ala Asp Ile Phe Gln Met				350
	355		360	
Gln Ala Thr Thr Arg Tyr Pro Gly Ala Tyr Tyr Ile Phe Gln Ile Lys				365
	370		375	
Ser Gly Asn Glu Gly Arg Glu Phe Tyr Met Arg Gln Thr Gly Pro Ile				380
385		390		395
Ser Ala Thr Leu Val Met Thr Arg Pro Ile Lys Gly Pro Arg Glu Ile				400
	405		410	
Gln Leu Asp Leu Glu Met Ile Thr Val Asn Thr Val Ile Asn Phe Arg				415
	420		425	
Gly Ser Ser Val Ile Arg Leu Arg Ile Tyr Val Ser Gln Tyr Pro Phe				430
	435		440	
			445	

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2550 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: CORNNOT01
- (B) CLONE: 45517

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCAAGATTGT TGTGAGGAGT CTAGCCAGTT GGTGAGCGCT GTAATCTGAA CCAGCTGTGT

60

```

CCAGACTGAG GCCCATTTTG CATTATTTAA CATACTTAGA AAATGAAGTG TTCATTTTAA 120
ACATTCCCTCC TCCAATTGGT TTAATGCTGA ATTACTGAAG AGGGCTAAGC AAAACCAGGT 180
GCTTGCGCTG AGGGCTCTGC AGTGCGTGGG AGGACCCCGG CGCTCTCCCC GTGTCTCTC 240
CACGACTCGC TCGGCCCCCTC TGAATAAAAA CACCCGCGAG CCCCAGGGGC CCAGAGGAGG 300
CCGACGTGCC CGAGCTCCTC CGGGGGTCCC GCGCGGAGC TTTCTTCTCG CTTTCGCATC 360
TCCTCTCTCG GCGTCTTGGA CATGCCAGGA ATAAAAAGGA TACTCACTGT TACCATTCTG 420
GCTCTCTGTC TTCCAAGCCC TGGGAATGCA CAGGCACAGT GCACGAATGG CTTTGACCTG 480
GATCGCCAGT CAGGACAGTG TTTAGATATT GATGAATGCC GAACCATCCC CGAGGCCTGC 540
CGAGGAGACA TGATGTGTGT TAACCAAAAT GCGGGGTATT TATGCATTCC CCGGACAAAC 600
CCTGTGTATC GAGGGCCCTA CTCGAACCCC TACTCGACCC CCTACTCAGG TCCGTACCCA 660
GCAGCTGCCC CACCACTCTC AGCTCCAAAC TATCCACGA TCTCCAGGCC TCTTATATGC 720
CGCTTTGGAT ACCAGATGGA TGAAAGCAAC CAATGTGTGG ATGTGGACGA GTGTGCAACA 780
GATTCCCAAC AGTGCAACCC CACCCAGATC TGCATCAATA CTGAAGGCGG GTACACCTGC 840
TCCTGCACCG ACGGATATTG GCTTCTGGAA GGCCAGTGCT TAGACATTGA TGAATGTCCG 900
TATGGTTACT GCCAGCAGCT CTGTGCGAAT GTTCTCGGAT CCTATTCTTG TACATGCAAC 960
CCTGGTTTTA CCTCAATGA GGATGGAAGG TCTTGCCAAG ATGTGAACGA GTGTGCCACC 1020
GAGAACCCCT GCGTGCAAAAC CTGCGTCAAC ACCTACGGCT CTTTCATCTG CCGCTGTGAC 1080
CCAGGATATG AACTTGAGGA AGATGCGTTC CATTGCAGTG ATATGGACGA GTGCAGCTTC 1140
TCTGAGTTCC TCTGCCAACA TGAGTGTGTG AACCAGCCCG GCACATACTT CTGCTCCTGC 1200
CCTCCAGGCT ACATCCTGCT GGATGACAAC CGAAGCTGCC AAGACATCAA CGAATGTGAG 1260
CACAGGAACC ACACGTGCAA CCTGCAGCAG ACSTGCTACA ATTTACAAGG GGGCTTCAAA 1320
TGATCGACC CCAATCCGCTG TGAGGAGCCT TATCTGAGGA TCAGTGATAA CCGCTGTATG 1380
TGTCCTGCTG AGAACCCCTG CTGCAGAGAC CAGCCCTTTA CCATCTTGTA CCGGGACATG 1440
GACGTGGTGT CAGGACGCTC CGTTCCCGCT GACATCTTCC AAATGCAAGC CACGACCCCG 1500
TACCCTGGGG CCTATTACAT TTTCAGATC AAATCTGGGA ATGAGGGCAG AGAATTTTAC 1560
ATGCGGCAAA CGGGCCCCAT CAGTGCCACC CTGGTGATGA CACGCCCAT CAAAGGGCCC 1620
CGGGAAATCC AGCTGGACTT GGAAATGATC ACTGTCAACA CTGTCAATCAA CTTCAGAGGC 1680
AGCTCCGTGA TCCGACTGCG GATATATGTG TCGCAGTACC CATTCTGAGC CTCGGGCTGG 1740
AGCCTCCGAG GCTGCCTCTC ATTGGCACCA AGGGACAGGA GAAGAGAGGA AATAACAGAG 1800
AGAATGAGAG CGACACAGAC GTTAGGCATT TCCTGCTGAA CGTTTCCCCG AAGAGTCAGC 1860
CCCGACTTCC TGACTCTCAC CTGTACTATT GCAGACCTGT CACCCTGCAG GACTTGCCAC 1920
CCCCAGTTCC TATGACACAG TTATCAAAAA GTATTATCAT TGCTCCCCCTG ATAGAAGATT 1980
GTTGGTGAAT TTTCAAGGCC TTCAGTTTAT TTCCACTATT TTCAAAGAAA ATAGATTAGG 2040
TTTGCGGGGG TCTGAGTCTA TGTTCAAAGC CTGTGAACAG CTTGCTGTCA CTTCTTCACC 2100
TCTTCCACTC CTTCTCTCAC TGTGTTACTG CTTTGCAAAG ACCCGGGAGC TGGCGGGGAA 2160
CCCTGGGAGT AGCTAGTTTG CTTTTTTCGT ACACAGAGAA GGCTATGTAA ACAAAACCACA 2220
GCAGGATCGA AGGGTTTTTA GAGAATGTGT TTCAAAACCA TGCTTGGTAT TTTCAACCAT 2280
AAAAAGAAGT TCAAGTTGTC TTAAATTTGT ATAACGGTTT AATTCTGTCT TGTCTATTTT 2340
GAGTATTTT AAAAATATG TCGTAGAATT CCTTCGAAAG GCCTTCAGAC ACATGCTATG 2400
TTCTGTCTTC CCAAACCCAG TCTCCTCTCC ATTTTAGCCC AGTGTTTTCT TTGAGGACCC 2460
CTTAATCTTG CTTTCTTAG AATTTTACC CAATTGGATT GGAATGCAGA GGTCTCCAAA 2520
CTGATTAAAT ATTTGAAGAG AAAAAAAAAA 2550

```

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 540 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT13
- (B) CLONE: 1621777

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Gly Thr Thr Ala Arg Ala Ala Leu Val Leu Thr Tyr Leu Ala Val
1           5           10           15
Ala Ser Ala Ala Ser Glu Gly Gly Phe Thr Ala Thr Gly Gln Arg Gln
20          25          30
Leu Arg Pro Glu His Phe Gln Glu Val Gly Tyr Ala Ala Pro Pro Ser
35          40          45
Pro Pro Leu Ser Arg Ser Leu Pro Met Asp His Pro Asp Ser Ser Gln
50          55          60

```



His	Gly	Pro	Pro	Phe	Glu	Gly	Gln	Ser	Gln	Val	Gln	Pro	Pro	Pro	Ser	65	70	75	80
Gln	Glu	Ala	Thr	Pro	Leu	Gln	Gln	Glu	Lys	Leu	Leu	Pro	Ala	Gln	Leu	85	90	95	
Pro	Ala	Glu	Lys	Glu	Val	Gly	Pro	Pro	Leu	Pro	Gln	Glu	Ala	Val	Pro	100	105	110	
Leu	Gln	Lys	Glu	Leu	Pro	Ser	Leu	Gln	His	Pro	Asn	Glu	Gln	Lys	Glu	115	120	125	
Gly	Met	Pro	Ala	Pro	Phe	Gly	Asp	Gln	Ser	His	Pro	Glu	Pro	Glu	Ser	130	135	140	
Trp	Asn	Ala	Ala	Gln	His	Cys	Gln	Gln	Asp	Arg	Ser	Gln	Gly	Gly	Trp	145	150	155	160
Gly	His	Arg	Leu	Asp	Gly	Phe	Pro	Pro	Gly	Arg	Pro	Ser	Pro	Asp	Asn	165	170	175	
Leu	Asn	Gln	Ile	Cys	Leu	Pro	Asn	Arg	Gln	His	Val	Val	Tyr	Gly	Pro	180	185	190	
Trp	Asn	Leu	Pro	Gln	Ser	Ser	Tyr	Ser	His	Leu	Thr	Arg	Gln	Gly	Glu	195	200	205	
Thr	Leu	Asn	Phe	Leu	Glu	Ile	Gly	Tyr	Ser	Arg	Cys	Cys	His	Cys	Arg	210	215	220	
Ser	His	Thr	Asn	Arg	Leu	Glu	Cys	Ala	Lys	Leu	Val	Trp	Glu	Glu	Ala	225	230	235	240
Met	Ser	Arg	Phe	Cys	Glu	Ala	Glu	Phe	Ser	Val	Lys	Thr	Arg	Pro	His	245	250	255	
Trp	Cys	Cys	Thr	Arg	Gln	Gly	Glu	Ala	Arg	Phe	Ser	Cys	Phe	Gln	Glu	260	265	270	
Glu	Ala	Pro	Gln	Pro	His	Tyr	Gln	Leu	Arg	Ala	Cys	Pro	Ser	His	Gln	275	280	285	
Pro	Asp	Ile	Ser	Ser	Gly	Leu	Glu	Leu	Pro	Phe	Pro	Pro	Gly	Val	Pro	290	295	300	
Thr	Leu	Asp	Asn	Ile	Lys	Asn	Ile	Cys	His	Leu	Arg	Arg	Phe	Arg	Ser	305	310	315	320
Val	Pro	Arg	Asn	Leu	Pro	Ala	Thr	Asp	Pro	Leu	Gln	Arg	Glu	Leu	Leu	325	330	335	
Ala	Leu	Ile	Gln	Leu	Glu	Arg	Glu	Phe	Gln	Arg	Cys	Cys	Arg	Gln	Gly	340	345	350	
Asn	Asn	His	Thr	Cys	Thr	Trp	Lys	Ala	Trp	Glu	Asp	Thr	Leu	Asp	Lys	355	360	365	
Tyr	Cys	Asp	Arg	Glu	Tyr	Ala	Val	Lys	Thr	His	His	His	Leu	Cys	Cys	370	375	380	
Arg	His	Pro	Pro	Ser	Pro	Thr	Arg	Asp	Glu	Cys	Phe	Ala	Arg	Arg	Ala	385	390	395	400
Pro	Tyr	Pro	Asn	Tyr	Asp	Arg	Asp	Ile	Leu	Thr	Ile	Asp	Ile	Gly	Arg	405	410	415	
Val	Thr	Pro	Asn	Leu	Met	Gly	His	Leu	Cys	Gly	Asn	Gln	Arg	Val	Leu	420	425	430	

Thr Lys His Lys His Ile Pro Gly Leu Ile His Asn Met Thr Ala Arg  
 435 440 445  
 Cys Cys Asp Leu Pro Phe Pro Glu Gln Ala Cys Cys Ala Glu Glu Glu  
 450 455 460  
 Lys Leu Thr Phe Ile Asn Asp Leu Cys Gly Pro Arg Arg Asn Ile Trp  
 465 470 475 480  
 Arg Asp Pro Ala Leu Cys Cys Tyr Leu Ser Pro Gly Asp Glu Gln Val  
 485 490 495  
 Asn Cys Phe Asn Ile Asn Tyr Leu Arg Asn Val Ala Leu Val Ser Gly  
 500 505 510  
 Asp Thr Glu Asn Ala Lys Gly Gln Gly Glu Gln Gly Ser Thr Gly Gly  
 515 520 525  
 Thr Asn Ile Ser Ser Thr Ser Glu Pro Lys Glu Glu  
 530 535 540

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1899 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT13  
 (B) CLONE: 162177

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGGGTGCAAG	CTCACAACCG	TAACAGCCAC	CAGACAAGCT	TCAGTGGCCG	GCCCTTCACA	60
TCCAGACTTG	CCTGAGAGGA	CCCACCTCTG	AGTGTCCAGT	GGTCAGTTGC	CCCAGGATGG	120
GGACCACAGC	CAGAGCAGCC	TTGGTCTTGA	CCTATTGTC	TGTTGCTTCT	GCTGCCTCTG	180
AGGGAGGCTT	CACGGCTACA	GGACAGAGGC	AGCTGAGGCC	AGAGCACTTT	CAAGAAGTTG	240
GCTACGCAGC	TCCCCCTCC	CCACCCTAT	CCCAGAGCCT	CCCCATGGAT	CACCCTGACT	300
CCTCTCAGCA	TGGCCCTCCC	TTTGAGGGAC	AGAGTCAAGT	GCAGCCCCCT	CCCTCTCAGG	360
AGGCCACCCC	TCTCCAACAG	GAAAAGCTGC	TACCTGCCCA	ACTCCCTGCT	GAAAAGGAAG	420
TGGGTCCCCC	TCTCCCTCAG	GAAGCTGTCC	CCCTCCAAA	AGAGCTGCCC	TCTCTCCAGC	480
ACCCCAATGA	ACAGAAGGAA	GGAATGCCAG	CTCCATTGG	GGACCAGAGC	CATCCAGAAC	540
CTGAGTCCTG	GAATGCAGCC	CAGCACTGCC	AACAGGACCG	GTCCCAAGGG	GGCTGGGGCC	600
ACCGGCTGGA	TGGCTTCCCC	CCTGGGCGGC	CTTCTCCAGA	CAATCTGAAC	CAAATCTGCC	660
TTCTTAACCG	TCAGCATGTG	GTATATGGTC	CCTGGAACCT	ACCACAGTCC	AGCTACTCCC	720
ACCTCACTCG	CCAGGGTGAG	ACCCTCAATT	TCCTGGAGAT	TGGATATTCC	CGCTGCTGCC	780
ACTGCCGCAG	CCACACAAAC	CGCCTAGAGT	GTGCCAACT	TGTGTGGGAG	GAAGCAATGA	840
GCCGATTCTG	TGAGGCCGAG	TTCTCGGTCA	AGACCCGACC	CCACTGGTGC	TGCACGCGGC	900
AGGGGGAGGC	TGGTTTCTCC	TGCTTCCAGG	AGGAAGCTCC	CCAGCCACAC	TACCAGCTCC	960
GGGCCTGCCC	CAGCCATCAG	CCTGATATTT	CCTCGGGTCT	TGAGCTGCCT	TTCCCTCTCTG	1020
GGGTGCCAC	ATTGGACAAT	ATCAAGAACA	TCTGCCACCT	GAGGCGCTTC	CGCTCTGTGC	1080
CACGCAACCT	GCCAGCTACT	GACCCCTTAC	AAAGGGAGCT	GCTGGCACTG	ATCCAGCTGG	1140
AGAGGGAGTT	CCAGCGCTGC	TGCCGCCAGG	GGAACAATCA	CACCTGTACA	TGGAAGGCCT	1200
GGGAGGATAC	CCTTGACAAA	TACTGTGACC	GGGAGTATGC	TGTGAAGACC	CACCACCACT	1260
TGTGTTGCCG	CCACCTCCC	AGCCCTACTC	GGGATGAGTG	CTTTGCCCGT	CGGGCTCCTT	1320
ACCCCAACTA	TGACCGGGAC	ATCTTGACCA	TTGACATCGG	TCGAGTCACC	CCCAACCTCA	1380
TGGGCCACCT	CTGTGGAAC	CAAAGAGTTC	TCACCAAGCA	TAAACATATT	CCTGGGCTGA	1440
TCCACAACAT	GACTGCCCGC	TGCTGTGACC	TGCCATTTCC	AGAACAGGCC	TGCTGTGCAG	1500
AGGAGGAGAA	ATTAACCTTC	ATCAATGATC	TGTGTGGTCC	CCGACGTAAC	ATCTGGCGAG	1560
ACCTTGCCCT	CTGCTGTTAC	CTGAGTCTTG	GGGATGAACA	GGTCAACTGC	TTCAACATCA	1620
ATTATCTGAG	GAACGTGGCT	CTAGTGTCTG	GAGACACTGA	GAACGCCAAG	GGCCAGGGGG	1680
AGCAGGGCTC	AACCTGGAGG	ACAAATATCA	GCTCCACCTC	TGAGCCCAAG	GAAGAATGAG	1740
TCACCCCAAG	GCCCTAGAGG	GTCAGATGGG	GGGAACCCCA	CCCTGCCCCA	CCCATCTGAA	1800
CACATCATAC	ACTAAACACC	TCTTGGATTT	GGTGTCTCTA	TTGTCTATCT	AATGTCTCAC	1860
CCGCAGTGTT	TTAAGTGGAT	CTTGGTGCCC	TGGCCGAGG			1899

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 387 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank  
 (B) CLONE: 458228

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

Met Ala Thr Ser Gly Val Leu Pro Gly Gly Gly Phe Val Ala Ser Ala
 1          5          10          15
Ala Ala Val Ala Gly Pro Glu Met Gln Thr Gly Arg Asn Asn Phe Val
 20          25          30
Ile Arg Arg Asn Pro Ala Asp Pro Gln Arg Ile Pro Ser Asn Pro Ser
 35          40          45
His Arg Ile Gln Cys Ala Ala Gly Tyr Glu Gln Ser Glu His Asn Val
 50          55          60
Cys Gln Asp Ile Asp Glu Cys Thr Ala Gly Thr His Asn Cys Arg Ala
 65          70          75          80
Asp Gln Val Cys Ile Asn Leu Arg Gly Ser Phe Ala Cys Gln Cys Pro
 85          90          95
Pro Gly Tyr Gln Lys Arg Gly Glu Gln Cys Val Asp Ile Asp Glu Cys
100          105          110
Thr Ile Pro Pro Tyr Cys His Gln Arg Cys Val Asn Thr Pro Gly Ser
115          120          125
Phe Tyr Cys Gln Cys Ser Pro Gly Phe Gln Leu Ala Ala Asn Asn Tyr
130          135          140
Thr Cys Val Asp Ile Asn Glu Cys Asp Ala Ser Asn Gln Cys Ala Gln
145          150          155          160
Gln Cys Tyr Asn Ile Leu Gly Ser Phe Ile Cys Gln Cys Asn Gln Gly
165          170          175
Tyr Glu Leu Ser Ser Asp Arg Leu Asn Cys Glu Asp Ile Asp Glu Cys
180          185          190
Arg Thr Ser Ser Tyr Leu Cys Gln Tyr Gln Cys Val Asn Glu Pro Gly
195          200          205
Lys Phe Ser Cys Met Cys Pro Gln Gly Tyr Gln Val Val Arg Ser Arg
210          215          220
Thr Cys Gln Asp Ile Asn Glu Cys Glu Thr Thr Asn Glu Cys Arg Glu
225          230          235          240
Asp Glu Met Cys Trp Asn Tyr His Gly Gly Phe Arg Cys Tyr Pro Arg
245          250          255
Asn Pro Cys Gln Asp Pro Tyr Ile Leu Thr Pro Glu Asn Arg Cys Val
260          265          270
Cys Pro Val Ser Asn Ala Met Cys Arg Glu Leu Pro Gln Ser Ile Val
275          280          285
Tyr Lys Tyr Met Ser Ile Arg Ser Asp Arg Ser Val Pro Ser Asp Ile
290          295          300
Phe Gln Ile Gln Ala Thr Thr Ile Tyr Ala Asn Thr Ile Asn Thr Phe
305          310          315          320
Arg Ile Lys Ser Gly Asn Glu Asn Gly Glu Phe Tyr Leu Arg Gln Thr
325          330          335
Ser Pro Val Ser Ala Met Leu Val Leu Val Lys Ser Leu Ser Gly Pro
340          345          350
Arg Glu His Ile Val Asp Leu Glu Met Leu Thr Val Ser Ser Ile Gly
355          360          365

```

Thr Phe Arg Thr Ser Ser Val Leu Arg Leu Thr Ile Ile Val Gly Pro  
 370 375 380  
 Phe Ser Phe  
 385

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 559 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 496120

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Thr Val Ser Arg Ala Ala Leu Ile Leu Ala Cys Leu Ala Leu  
 1 5 10 15  
 Ala Ser Ala Ala Ser Glu Gly Ala Phe Lys Ala Ser Asp Gln Arg Glu  
 20 25 30  
 Met Thr Pro Glu Arg Leu Phe Gln His Leu His Glu Val Gly Tyr Ala  
 35 40 45  
 Ala Pro Pro Ser Leu Pro Gln Thr Arg Arg Leu Arg Val Asp His Ser  
 50 55 60  
 Val Thr Ser Leu His Asp Pro Pro Leu Phe Glu Glu Gln Arg Glu Val  
 65 70 75 80  
 Gln Pro Pro Ser Ser Pro Glu Asp Ile Pro Val Tyr Glu Glu Asp Trp  
 85 90 95  
 Pro Thr Phe Leu Asn Pro Asn Val Asp Lys Ala Gly Pro Ala Val Pro  
 100 105 110  
 Gln Glu Ala Ile Pro Leu Gln Lys Glu Gln Pro Pro Pro Gln Val His  
 115 120 125  
 Ile Glu Gln Lys Glu Ile Asp Pro Pro Ala Gln Pro Gln Glu Glu Ile  
 130 135 140  
 Val Gln Lys Glu Val Lys Pro His Thr Leu Ala Gly Gln Leu Pro Pro  
 145 150 155 160  
 Glu Pro Arg Thr Trp Asn Pro Ala Arg His Cys Gln Gln Gly Arg Arg  
 165 170 175  
 Gly Val Trp Gly His Arg Leu Asp Gly Phe Pro Pro Gly Arg Pro Ser  
 180 185 190  
 Pro Asp Asn Leu Lys Gln Ile Cys Leu Pro Glu Arg Gln His Val Ile  
 195 200 205  
 Tyr Gly Pro Trp Asn Leu Pro Gln Thr Gly Tyr Ser His Leu Ser Arg  
 210 215 220  
 Gln Gly Glu Thr Leu Asn Val Leu Glu Thr Gly Tyr Ser Arg Cys Cys  
 225 230 235 240  
 Pro Cys Arg Ser Asp Thr Asn Arg Leu Asp Cys Leu Lys Leu Val Trp  
 245 250 255  
 Glu Asp Ala Met Thr Gln Phe Cys Glu Ala Glu Phe Ser Val Lys Thr  
 260 265 270  
 Arg Pro His Leu Cys Cys Arg Leu Arg Gly Glu Glu Arg Phe Ser Cys  
 275 280 285  
 Phe Gln Lys Glu Ala Pro Arg Pro Asp Tyr Leu Leu Arg Pro Cys Pro  
 290 295 300  
 Val His Gln Asn Gly Met Ser Ser Gly Pro Gln Leu Pro Phe Pro Pro  
 305 310 315 320  
 Gly Leu Pro Thr Pro Asp Asn Val Lys Asn Ile Cys Leu Leu Arg Arg  
 325 330 335  
 Phe Arg Ala Val Pro Arg Asn Leu Pro Ala Thr Asp Ala Ile Gln Arg  
 340 345 350  
 Gln Leu Gln Ala Leu Thr Arg Leu Glu Thr Glu Phe Gln Arg Cys Cys  
 355 360 365

Arg	Gln	Gly	His	Asn	His	Thr	Cys	Thr	Trp	Lys	Ala	Trp	Glu	Gly	Thr
	370					375					380				
Leu	Asp	Gly	Tyr	Cys	Glu	Arg	Glu	Leu	Ala	Ile	Lys	Thr	His	Pro	His
385					390					395					400
Ser	Cys	Cys	His	Tyr	Pro	Pro	Ser	Pro	Ala	Arg	Asp	Glu	Cys	Phe	Ala
				405					410					415	
His	Leu	Ala	Pro	Tyr	Pro	Asn	Tyr	Asp	Arg	Asp	Ile	Leu	Thr	Leu	Asp
			420					425					430		
Leu	Ser	Arg	Val	Thr	Pro	Asn	Leu	Met	Gly	Gln	Leu	Cys	Gly	Ser	Gly
			435				440					445			
Arg	Val	Leu	Ser	Lys	His	Lys	Gln	Ile	Pro	Gly	Leu	Ile	Gln	Asn	Met
	450					455					460				
Thr	Val	Arg	Cys	Cys	Glu	Leu	Pro	Tyr	Pro	Glu	Gln	Ala	Cys	Cys	Gly
465					470					475					480
Glu	Glu	Glu	Lys	Leu	Ala	Phe	Ile	Glu	Asn	Leu	Cys	Gly	Pro	Arg	Arg
				485					490					495	
Asn	Ser	Trp	Lys	Asp	Pro	Ala	Leu	Cys	Cys	Asp	Leu	Ser	Pro	Glu	Asp
			500					505					510		
Lys	Gln	Ile	Asn	Cys	Phe	Asn	Thr	Asn	Tyr	Leu	Arg	Asn	Val	Ala	Leu
	515						520					525			
Val	Ala	Gly	Asp	Thr	Gly	Asn	Ala	Thr	Gly	Leu	Gly	Glu	Gln	Gly	Pro
	530					535					540				
Thr	Arg	Gly	Thr	Asp	Ala	Asn	Pro	Ala	Pro	Gly	Ser	Lys	Glu	Glu	
545					550					555					

What is claimed is:

1. An isolated and purified polynucleotide encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:1.

5

2. An isolated and purified polynucleotide which is fully complementary to the polynucleotide of claim 1.

3. An isolated and purified polynucleotide consisting of SEQ ID NO:2.

10

4. An isolated and purified polynucleotide which is fully complementary to the polynucleotide sequence of claim 3.

5. An expression vector containing the polynucleotide of claim 1.

15

6. A host cell containing the expression vector of claim 5.

7. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, the method comprising the steps of:

20

a) culturing the host cell of claim 6 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

1/17

9	18	27	36	45	54
CCA AGA TTG TTG TGA GGA GTC TAG CCA GTT GGT GAG CGC TGT AAT CTG AAC CAG					
63	72	81	90	99	108
CTG TGT CCA GAC TGA GGC CCC ATT TGC ATT ATT TAA CAT ACT TAG AAA ATG AAG					
117	126	135	144	153	162
TGT TCA TTT TTA ACA TTC CTC CTC CAA TTG GTT TAA TGC TGA ATT ACT GAA GAG					
171	180	189	198	207	216
GGC TAA GCA AAA CCA GGT GGT TGC GCT GAG GGC TCT GCA GTG GCT GGG AGG ACC					
225	234	243	252	261	270
CCG GCG CTC TCC CCG TGT CCT CTC CAC GAC TCG CTC GGC CCC TCT GGA ATA AAA					
279	288	297	306	315	324
CAC CCG CGA GCC CCG AGG GCC CAG AGG AGG CCG TGC TGC CCG AGC TCC TCC GGG					
333	342	351	360	369	378
GGT CCC GCC CGC GAG CTT TCT TCT CGC CTT CGC ATC TCC TCC TCG CGC GTC TTG					
387	396	405	414	423	432
GAC ATG CCA GGA ATA AAA AGG ATA CTC ACT GTT ACC ATT CTG GCT CTC TGT CTT					
M P G I K R I L T V T I L A L C L					

FIGURE 1A

441	CCA AGC CCT GGG AAT GCA CAG GCA CAG TGC ACG AAT GGC TTT GAC CTG GAT CGC	468	477	486
P S P G N A Q A Q A Q C T N G F D L D R				
495	CAG TCA GGA CAG TGT TTA GAT ATT GAT GAA TGC CGA ACC ATC CCC GAG GCC TGC	522	531	540
Q S G Q C L D I D E C R T I P E A C				
549	CGA GGA GAC ATG ATG TGT GTT AAC CAA AAT GGC GGG TAT TTA TGC ATT CCC CGG	576	585	594
R G D M M C V N Q N G G Y L C I P R				
603	ACA AAC CCT GTG TAT CGA GGG CCC TAC TCG AAC CCC TAC TCG ACC CCC TAC TCA	630	639	648
T N P V Y R G G P Y S N P Y S T P Y S				
657	GGT CCG TAC CCA GCA GCT GCC CCA CTC TCA GCT CCA AAC TAT CCC ACG ATC	684	693	702
G P Y P A A A P P L S A P N Y P T I				
711	TCC AGG CCT CTT ATA TGC CGC TTT GGA TAC CAG ATG GAT GAA AGC AAC CAA TGT	738	747	756
S R P L I C R F G Y Q M D E S N Q C				
765	GTG GAT GTG GAC GAG TGT GCA ACA GAT TCC CAC CAG TGC AAC CCC ACC CAG ATC	792	801	810
V D V D E C A T D S H Q C N P T Q I				

FIGURE 1B



819	828	837	846	855	864
TGC ATC AAT	ACT GAA GGC	TAC ACC TGC	TCC TGC ACC	GAC GGA TAT	TGG CTT
C I N T	E G G Y	T C C S	C T D G	Y W L	
873	882	891	900	909	918
CTG GAA GGC	CAG TGC TTA	GAC ATT GAT	GAA TGT CGC	TAT TGC	CAG CAG
L E G Q	C L D I	D E C R	Y G Y C	Q Q	
927	936	945	954	963	972
CTC TGT GCG	AAT GTT CCT	GGA TCC TAT	TCT TGT ACA	TGC AAC	CCT GGT
L C A N	V P G S	Y S C T	C N P G	F T	
981	990	999	1008	1017	1026
CTC AAT GAG	GAT GGA AGG	TCT TGC CAA	GAT GTG AAC	GAG TGT	GCC ACC
L N E D	G R S C	Q D V N	E C A T	E N	
1035	1044	1053	1062	1071	1080
CCC TGC GTG	CAA ACC TGC	GTC AAC ACC	TAC GGC TCT	TTC ATC	TGC CGC
P C V Q	T C V N	T Y G S	F I C R	C D	
1089	1098	1107	1116	1125	1134
CCA GGA TAT	GAA CTT GAG	GAA GAT GGC	GTT CAT TGC	AGT GAT	ATG GAC
P G Y E	L E E D	G V H C	S D M D	E C	
1143	1152	1161	1170	1179	1188
AGC TTC TCT	GAG TTC CTC	TGC CAA CAT	GAG TGT GTG	AAC CAG	CCC GGC
S F S E	F L C Q	H E C V	N Q P G	T Y	

FIGURE 1C

1197	1206	1215	1224	1233	1242
TTC TGC TCC TGC CCT CCA GGC TAC ATC CTG CTG GAT GAC AAC CGA AGC TGC CAA					
F C S C P P G Y I L L D D N R S C Q					
1251	1260	1269	1278	1287	1296
GAC ATC AAC GAA TGT GAG CAC AGG AAC CAC ACG TGC AAC CTG CAG CAG ACG TGC					
D I N E C E H R N H T C N L Q Q T C					
1305	1314	1323	1332	1341	1350
TAC AAT TTA CAA GGG GGC TTC AAA TGC ATC GAC CCC ATC CGC TGT GAG GAG CCT					
Y N L Q G G F K C I D P I R C E E P					
1359	1368	1377	1386	1395	1404
TAT CTG AGG ATC AGT GAT AAC CGC TGT ATG TGT CCT GCT GAG AAC CCT GGC TGC					
Y L R I S D N R C M C P A E N P G C					
1413	1422	1431	1440	1449	1458
AGA GAC CAG CCC TTT ACC ATC TTG TAC CGG GAC ATG GAC GTG GTG TCA GGA CGC					
R D Q P F T I L Y R D M D V V S G R					
1467	1476	1485	1494	1503	1512
TCC GTT CCC GCT GAC ATC TTC CAA ATG CAA GCC ACG ACC CGC TAC CCT GGC GCC					
S V P A D I F Q M Q A T T R Y P G A					
1521	1530	1539	1548	1557	1566
TAT TAC ATT TTC CAG ATC AAA TCT GGG AAT GAG GGC AGA GAA TTT TAC ATG CGG					
Y Y I F Q I K S G N E G R E F Y M R					

FIGURE ID

5/17

1575	1584	1593	1602	1611	1620
CAA ACG GGC CCC ATC AGT GCC ACC CTG GTG ATG ACA CGC CCC ATC AAA GGG CCC					
Q T G P I S A T L V M T R P I K G P					
1629	1638	1647	1656	1665	1674
CGG GAA ATC CAG CTG GAC TTG GAA ATG ATC ACT GTC AAC ACT GTC ATC AAC TTC					
R E I Q L D L E M I T V N T V I N F					
1683	1692	1701	1710	1719	1728
AGA GGC AGC TCC GTG ATC CGA CTG CGG ATA TAT GTG TCG CAG TAC CCA TTC TGA					
R G S S V I R L R I Y V S Q Y P F					
1737	1746	1755	1764	1773	1782
GCC TCG GGC TGG AGC CTC CGA CGC TGC CTC TCA TTG GCA CCA AGG GAC AGG AGA					
1791	1800	1809	1818	1827	1836
AGA GAG GAA ATA ACA GAG AGA ATG AGA GCG ACA CAG ACG TTA GGC ATT TCC TGC					
1845	1854	1863	1872	1881	1890
TGA ACG TTT CCC CGA AGA GTC AGC CCC GAC TTC CTG ACT CTC ACC TGT ACT ATT					
1899	1908	1917	1926	1935	1944
GCA GAC CTG TCA CCC TGC AGG ACT TGC CAC CCC CAG TTC CTA TGA CAC AGT TAT					
1953	1962	1971	1980	1989	1998
CAA AAA GTA TTA TCA TTG CTC CCC TGA TAG AAG ATT GTT GGT GAA TTT TCA AGG					

FIGURE 1E

6/17

2007	2016	2025	2034	2043	2052
CCT TCA GTT TAT TTC CAC TAT TTT CAA AGA AAA TAG ATT AGG TTT GCG GGC GTC					
2061	2070	2079	2088	2097	2106
TGA GTC TAT GTT CAA AGA CTG TGA ACA GCT TGC TGT CAC TTC TTC ACC TCT TCC					
2115	2124	2133	2142	2151	2160
ACT CCT TCT CTC ACT GTG TTA CTG CTT TGC AAA GAC CCG GGA GCT GGC GGC GAA					
2169	2178	2187	2196	2205	2214
CCC TGG GAG TAG CTA GTT TGC TTT TTG CGT ACA CAG AGA AGG CTA TGT AAA CAA					
2223	2232	2241	2250	2259	2268
ACC ACA GCA GGA TCG AAG GGT TTT TAG AGA ATG TGT TTC AAA ACC ATG CCT GGT					
2277	2286	2295	2304	2313	2322
ATT TTC AAC CAT AAA AGA AGT TTC AGT TGT CCT TAA ATT TGT ATA ACG GTT TAA					
2331	2340	2349	2358	2367	2376
TTC TGT CTT GTT CAT TTT GAG TAT TTT TAA AAA ATA TGT CGT AGA ATT CCT TCG					
2385	2394	2403	2412	2421	2430
AAA GGC CTT CAG ACA CAT GCT ATG TTC TGT CTT CCC AAA CCC AGT CTC CTC TCC					
2439	2448	2457	2466	2475	2484
ATT TTA GCC CAG TGT TTT CTT TGA GGA CCC CTT AAT CTT GCT TTC TTT AGA ATT					

FIGURE 1F

7/17

2493 2502 2511 2520 2529 2538  
TTT ACC CAA TTG GAT TGG AAT GCA GAG GTC TCC AAA CTG ATT AAA TAT TTG AAG

2547  
AGA AAA AAA AAA

FIGURE 1G

8/17

9	18	27	36	45	54
TG GGT GCA AGC TCA CAA CCG TAA CAG CCA CCA GAC AAG CTT CAG TGG CCG GCC					
63	72	81	90	99	108
CTT CAC ATC CAG ACT TGC CTG AGA GGA CCC ACC TCT GAG TGT CCA GTG GTC AGT					
117	126	135	144	153	162
TGC CCC AGG ATG GGG ACC ACA GCC AGA GCA GCC TTG GTC TTG ACC TAT TTG GCT					
M G T T A R A A A L V L Y L A					
171	180	189	198	207	216
GTT GCT TCT GCT GCC TCT GAG GGA GGC TTC ACG GCT ACA GGA CAG AGG CAG CTG					
V A S A A S E G G G F T A T G G Q R Q L					
225	234	243	252	261	270
AGG CCA GAG CAC TTT CAA GAA GTT GGC TAC GCA GCT CCC CCC TCC CCA CCC CTA					
R P E H F Q E V G Y A A P P S P P L					
279	288	297	306	315	324
TCC CGA AGC CTC CCC ATG GAT CAC CCT GAC TCC TCT CAG CAT GGC CCT CCC TTT					
S R S L P M D H P D S S S Q H G P P F					
333	342	351	360	369	378
GAG GGA CAG AGT CAA GTG CAG CCC CCT CCC TCT CAG GAG GCC ACC CCT CTC CAA					
E G Q S Q V Q Q P P P S Q Q E A T P L Q					

FIGURE 2A

9/17

387	CAG GAA AAG CTG CTA CCT GCT GAA AAG GAA GTG GGT CCC CCT	414	423	432
396	Q E K L L P A Q L P A E K E V G P P			
441	CTC CCT CAG GAA GCT GTC CCC CTC CAA AAA GAG CTG CCC TCT CTC CAG CAC CCC	468	477	486
450	L P Q E A V P P L Q K E L P S L Q H P			
495	AAT GAA CAG AAG GAA GGA ATG CCA GCT CCA TTT GGG GAC CAG AGC CAT CCA GAA	522	531	540
504	N E Q K E G M P A P F G D Q S H P E			
549	CCT GAG TCC TGG AAT GCA GCC CAG CAC TGC CAA CAG GAC CGG TCC CAA GGG GGC	576	585	594
558	P E S W N A A Q H C Q Q D R S Q G G			
603	TGG GGC CAC CGG CTG GAT GGC TTC CCC CCT GGG CGG CCT TCT CCA GAC AAT CTG	630	639	648
612	W G H R L D G F P P G R P S P D N L			
657	AAC CAA ATC TGC CTT CCT AAC CGT CAG CAT GTG GTA TAT GGT CCC TGG AAC CTA	684	693	702
666	N Q I C L P N R Q H V V Y G P W N L			
711	CCA CAG TCC AGC TAC TCC CAC CTC ACT CGC CAG GGT GAG ACC CTC AAT TTC CTG	738	747	756
720	P Q S S Y S H L T R Q G E T L N F L			

FIGURE 2B

10/17

765	GAG ATT GGA TAT TCC CGC TGC TGC CAC TGC CGC AGC CAC ACA AAC CGC CTA GAG	774	783	792	801	810
E I G Y S R C C H C R S H T N R L E						
819	TGT GCC AAA CTT GTG TGG GAG GAA GCA ATG AGC CGA TTC TGT GAG GCC GAG TTC	828	837	846	855	864
C A K L V W E A M S R F C A E F						
873	TCG GTC AAG ACC CGA CCC CAC TGG TGC TGC ACG CGG CAG GGG GAG GCT CGG TTC	882	891	900	909	918
S V K T R P H W C C T R Q G E A R F						
927	TCC TGC TTC CAG GAG GAA GCT CCC CAG CCA CAC TAC CAG CTC CGG GCC TGC CCC	936	945	954	963	972
S C F Q E E A P Q P H Y Q L R A C P						
981	AGC CAT CAG CCT GAT ATT TCC TCG GGT CTT GAG CTG CCT TTC CCT CCT GGG GTG	990	999	1008	1017	1026
S H Q P D I S S G L E L P F P G V						
1035	CCC ACA TTG GAC AAT ATC AAG AAC ATC TGC CAC CTG AGG CGC TTC CGC TCT GTG	1044	1053	1062	1071	1080
P T L D N I K N I C H L R R F S V						
1089	CCA CGC AAC CTG CCA GCT ACT GAC CCC CTA CAA AGG GAG CTG CTG GCA CTG ATC	1098	1107	1116	1125	1134
P R N L P A T D P L Q R E L L A L I						

FIGURE 2C



11/17

1143	1152	1161	1170	1179	1188
CAG CTG GAG AGG GAG TTC CAG CGC TGC TGC CGC CAG GGG AAC AAT CAC ACC TGT					
Q L E R E F Q R C C R Q G N N H T C					
1197	1206	1215	1224	1233	1242
ACA TGG AAG GCC TGG GAG GAT ACC CTT GAC AAA TAC TGT GAC CGG GAG TAT GCT					
T W K A W E E D T L D K Y C D R E Y A					
1251	1260	1269	1278	1287	1296
GTG AAG ACC CAC CAC CAC TTG TGT TGC CGC CAC CCT CCC AGC CCT ACT CGG GAT					
V K T H H H L C C R H P P S P T R D					
1305	1314	1323	1332	1341	1350
GAG TGC TTT GCC CGT CGG GCT CCT TAC CCC AAC TAT GAC CGG GAC ATC TTG ACC					
E C F A R R A P Y P N Y D R D I L T					
1359	1368	1377	1386	1395	1404
ATT GAC ATC GGT CGA GTC ACC CCC AAC CTC ATG GGC CAC CTC TGT GGA AAC CAA					
I D I G R V T P N L M G H L C G N Q					
1413	1422	1431	1440	1449	1458
AGA GTT CTC ACC AAG CAT AAA CAT ATT CCT GGG CTG ATC CAC AAC ATG ACT GCC					
R V L T K H K H I P G L I H N M T A					
1467	1476	1485	1494	1503	1512
CGC TGC TGT GAC CTG CCA TTT CCA GAA CAG GCC TGC TGT GCA GAG GAG GAG AAA					
R C C D L P F P E Q A C C A E E E K					

FIGURE 2D

12/17

1521 1530 1539 1548 1557 1566  
TTA ACC TTC ATC AAT GAT CTG TGT GGT CCC CGA CGT AAC ATC TGG CGA GAC CCT  
L T F I N D L C G P R R N I W R D P

1575 1584 1593 1602 1611 1620  
GCC CTC TGC TGT TAC CTG AGT CCT GGG GAT GAA CAG GTC AAC TGC TTC AAC ATC  
A L C C Y L S P G D E Q V N C F N I

1629 1638 1647 1656 1665 1674  
AAT TAT CTG AGG AAC GTG GCT CTA GTG TCT GGA GAC ACT GAG AAC GCC AAG GGC  
N Y L R N V A L V S G D T E N A K G

1683 1692 1701 1710 1719 1728  
CAG GGG GAG CAG GGC TCA ACT GGA GGA ACA AAT ATC AGC TCC ACC TCT GAG CCC  
Q G E Q G S T G G T N I S S T S E P

1737 1746 1755 1764 1773 1782  
AAG GAA GAA TGA GTC ACC CCA GAG CCC TAG AGG GTC AGA TGG GGG GAA CCC CAC  
K E E

1791 1800 1809 1818 1827 1836  
CCT GCC CCA CCC ATC TGA ACA CTC ATT ACA CTA AAC ACC TCT TGG ATT TGG TGT

1845 1854 1863 1872 1881 1890  
CCT CAT TGT CTA TCT AAT GTC TCA CCC GCA GTG TTT TAA GTG GAT CTT GGT GCC

1899  
CTG GCC CAG G

FIGURE 2E

13/17

1	M P G I K R I L T V T I L A L C L P S P	45517
1	M A T S G V L P G G G F V A S A A A V A	GI 458228
21	G N A Q A Q C T N G F D L D R Q S G Q C	45517
21	G P E M Q T G R N N F V I R R N P A D -	GI 458228
41	L D I D E C R T I P E A C R G D M M C V	45517
40	- - - - -	GI 458228
61	N Q N G G Y L C I P R T N P V Y R G P Y	45517
40	- - - - -	GI 458228
81	S N P Y S T P Y S G P Y P A A P P L S	45517
40	- - - - - P Q R	GI 458228
101	A P N Y P T I S R P L I C R F G Y Q M D	45517
43	I P S N P - - S H R I Q C A A G Y E Q S	GI 458228
121	E S N Q C V D V D E C A T D S H Q C N P	45517
61	E H N V C Q D I D E C T A G T H N C R A	GI 458228
141	T Q I C I N T E G G Y T C S C T D G Y W	45517
81	D Q V C I N L R G S F A C Q C P P G Y Q	GI 458228
161	L L E G Q C L D I D E C R Y G - Y C Q Q	45517
101	K R G E Q C V D I D E C T I P P Y C H Q	GI 458228
180	L C A N V P G S Y S C T C N P G F T L N	45517
121	R C V N T P G S F Y C Q C S P G F Q L A	GI 458228
200	E D G R S C Q D V N E C A T E N P C V Q	45517
141	A N N Y T C V D I N E C D A S N Q C A Q	GI 458228
220	T C V N T Y G S F I C R C D P G Y E L E	45517
161	Q C Y N I L G S F I C Q C N Q G Y E L S	GI 458228

FIGURE 3A

14/17

240	E D G V H C S D M D E C S F S E F L C Q	45517
181	S D R L N C E D I D E C R T S S Y L C Q	GI 458228
260	H E C V N Q P G T Y F C S C P P G Y I L	45517
201	Y Q C V N E P G K F S C M C P Q G Y Q V	GI 458228
280	L D D N R S C Q D I N E C E H R N H T C	45517
221	V R - S R T C Q D I N E C E T T N E - C	GI 458228
300	N L Q Q T C Y N L Q G G F K C I D P I R	45517
239	R E D E M C W N Y H G G F R C Y P R N P	GI 458228
320	C E E P Y L R I S D N R C M C P A E N P	45517
259	C Q D P Y I L T P E N R C V C P V S N A	GI 458228
340	G C R D Q P F T I L Y R D M D V V S G R	45517
279	M C R E L P Q S I V Y K Y M S I R S D R	GI 458228
360	S V P A D I F Q M Q A T T R Y P G A Y Y	45517
299	S V P S D I F Q I Q A T T I Y A N T I N	GI 458228
380	I F Q I K S G N E G R E F Y M R Q T G P	45517
319	T F R I K S G N E N G E F Y L R Q T S P	GI 458228
400	I S A T L V M T R P I K G P R E I Q L D	45517
339	V S A M L V L V K S L S G P R E H I V D	GI 458228
420	L E M I T V N T V I N F R G S S V I R L	45517
359	L E M L T V S S I G T F R T S S V L R L	GI 458228
440	R I Y V S Q Y P F	45517
379	T I I V G P F S F	GI 458228

FIGURE 3B

15/17

1	M G T T A R A A L V L T Y L A V A S A A	1621777
1	M G T V S R A A L I L A C L A L A S A A	GI 496120
21	S E G G F T A T G Q R Q L R P E - - - -	1621777
21	S E G A F K A S D Q R E M T P E R L F Q	GI 496120
37	H F Q E V G Y A A P P S P P L S R S L P	1621777
41	H L H E V G Y A A P P S L P Q T R R L R	GI 496120
57	M D H P D S S Q H G P P - F E G Q S Q V	1621777
61	V D H S V T S L H D P P L F E E Q R E V	GI 496120
76	Q P P P S Q E A T P L Q Q E K L L P A Q	1621777
81	Q P P S S P E D I P V Y E E D W P T F L	GI 496120
96	L P A E K E V G P P L P Q E A V P L Q K	1621777
101	N P N V D K A G P A V P Q E A I P L Q K	GI 496120
116	E L P S L Q - - - - - - - - - - H P	1621777
121	E Q P P P Q V H I E Q K E I D P P A Q P	GI 496120
124	N E - - - Q K E G M P A P F G D Q S H P	1621777
141	Q E E I V Q K E V K P H T L A G Q L P P	GI 496120
141	E P E S W N A A Q H C Q Q D R S Q G G W	1621777
161	E P R T W N P A R H C Q Q G R - R G V W	GI 496120
161	G H R L D G F P P G R P S P D N L N Q I	1621777
180	G H R L D G F P P G R P S P D N L K Q I	GI 496120
181	C L P N R Q H V V Y G P W N L P Q S S Y	1621777
200	C L P E R Q H V I Y G P W N L P Q T G Y	GI 496120
201	S H L T R Q G E T L N F L E I G Y S R C	1621777
220	S H L S R Q G E T L N V L E T G Y S R C	GI 496120

FIGURE 4A

16/17

221	C H C R S H T N R L E C A K L V W E E A	1621777
240	C P C R S D T N R L D C L K L V W E D A	GI 496120
241	M S R F C E A E F S V K T R P H W C C T	1621777
260	M T Q F C E A E F S V K T R P H L C C R	GI 496120
261	R Q G E A R F S C F Q E E A P Q P H Y Q	1621777
280	L R G E E R F S C F Q K E A P R P D Y L	GI 496120
281	L R A C P S H Q P D I S S G L E L P F P	1621777
300	L R P C P V H Q N G M S S G P Q L P F P	GI 496120
301	P G V P T L D N I K N I C H L R R F R S	1621777
320	P G L P T P D N V K N I C L L R R F R A	GI 496120
321	V P R N L P A T D P L Q R E L L A L I Q	1621777
340	V P R N L P A T D A I Q R Q L Q A L T R	GI 496120
341	L E R E F Q R C C R Q G N N H T C T W K	1621777
360	L E T E F Q R C C R Q G H N H T C T W K	GI 496120
361	A W E D T L D K Y C D R E Y A V K T H H	1621777
380	A W E G T L D G Y C E R E L A I K T H P	GI 496120
381	H L C C R H P P S P T R D E C F A R R A	1621777
400	H S C C H Y P P S P A R D E C F A H L A	GI 496120
401	P Y P N Y D R D I L T I D I G R V T P N	1621777
420	P Y P N Y D R D I L T L D L S R V T P N	GI 496120
421	L M G H L C G N Q R V L T K H K H I P G	1621777
440	L M G Q L C G S G R V L S K H K Q I P G	GI 496120
441	L I H N M T A R C C D L P F P E Q A C C	1621777
460	L I Q N M T V R C C E L P Y P E Q A C C	GI 496120

FIGURE 4B

17/17

461	A	E	E	E	K	L	T	F	I	N	D	L	C	G	P	R	R	N	I	W	1621777
480	G	E	E	E	K	L	A	F	I	E	N	L	C	G	P	R	R	N	S	W	GI 496120
481	R	D	P	A	L	C	C	Y	L	S	P	G	D	E	Q	V	N	C	F	N	1621777
500	K	D	P	A	L	C	C	D	L	S	P	E	D	K	Q	I	N	C	F	N	GI 496120
501	I	N	Y	L	R	N	V	A	L	V	S	G	D	T	E	N	A	K	G	Q	1621777
520	T	N	Y	L	R	N	V	A	L	V	A	G	D	T	G	N	A	T	G	L	GI 496120
521	G	E	Q	G	S	T	G	G	T	N	I	S	S	T	S	E	P	K	E	E	1621777
540	G	E	Q	G	P	T	R	G	T	D	A	N	P	A	P	G	S	K	E	E	GI 496120

FIGURE 4C